

KamLAND Data Analysis Production

B.E. Berger¹, M.P. Decowski², D.A. Dwyer², S.J. Freedman², B.K. Fujikawa¹, K.M. Heeger¹, L.L. Hsu¹, K.-B. Luk², H. Murayama², T.M. O'Donnell², H.M. Steiner², and L.A. Winslow² for the KamLAND collaboration

¹*Nuclear Science Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720*

²*Physics Department, University of California at Berkeley, Berkeley, California 94720*

In recent years the KamLAND experiment has not only provided strong evidence that neutrinos oscillate, but has made key measurements of fundamental standard model parameters of neutrino physics [1, 2]. The obstacle of most neutrino oscillation experiments is the low neutrino event rate relative to detector backgrounds. In KamLAND we detect approximately 30 events per second, but expect less than one per million to be a reactor neutrino interaction. Not only does this provide a challenge in the analysis of the data, but it is also a procedural problem to archive and process this large volume of data. Over the past year an effort has been made to construct a procedure to handle these challenges. The system will then be used in the production of the data for the next KamLAND reactor neutrino analysis. This will include more than a doubling of the total dataset and should give the most precise measurement of Δm_{12}^2 to date.

The data analysis production system was designed with the following goals:

Simple All collaborators should be able to easily participate in the production of KamLAND data.

Robust It should be easy to diagnose and recover data which fails to process correctly.

Summarized There should be a simple interface to summarize the current state of production and to document where data is stored.

The KamLAND detector currently generates roughly 200 GB of data per day. This data is transported weekly from Japan on magnetic tape and then archived in the High Performance Storage System (HPSS) at the National Energy Research Scientific Computing Center (NERSC) [3]. During production, this data is read out of HPSS and processed on the Parallel Distributed Systems Facility (PDSF) computer cluster. Since the last analysis of KamLAND data, over 400 additional days of data have been taken, bringing the total to almost 900 days.

To achieve the production goals outlined above, a system was constructed consisting of three parts: production scripts, database, and web interface. The production scripts are a robust set of computer programs which handle the submission and monitoring of jobs on PDSF. The database documents the status of all production jobs and the location of all data files.

The web interface provides simple and human-readable access to the data in the database (Fig. 1.)

This entire system has been thoroughly tested. Furthermore, a significant fraction of the collaboration has been trained in its use and have all submitted data for production. This initial data is undergoing a thorough physics study to

Statistics of all files

File Type	Number of files	Total Space (GB)
kdf	183527	43655
kdf-fme-index	1	0
kdf-index	1465	143
kdf-trigger	4627	320
kdf-trigger-index	4057	30
kdfz	309969	62565
kdfz-fme	72355	14461
kdfz-fme-index	2437	19
kdfz-index	2585	688
RECON	3942	86
sf	127976	30378
sf-end-of-run	1270	0
sfz	257831	60536
sfz-built	2114	0
TQ	358	175
Total:	974514	213055

Refresh

FIG. 1: Example display of file statistics from user interface.

evaluate the new KamLAND analysis tools. Once this is complete, the full production will begin in earnest. The entire production is expected to be completed in two months according to current estimates.

[1] K. Eguchi et al., Phys. Rev. Lett. **90**, 021802 (2003).

[2] T. Araki et al., Phys. Rev. Lett. **94**, 081801 (2005).

[3] URL <http://www.nersc.gov>.